Variation in *PNPLA3* is associated with outcomes in alcoholic liver disease

Chao Tian[1], Renee P. Stokowski[1], David Kershenobich[2], Dennis G. Ballinger[1,3], David A. Hinds[1]

1. Perlegen, 2021 Stierlin Court, Mountain View, CA 94043, USA.
2. Departamento de Medicina Experimental, Facultad de Medicina de la Universidad Nacional Autónoma de México, Mexico D.F.
3. Current address: Complete Genomics, 2071 Stierlin Court, Mountain View, CA 94043, USA.

Correspondence should be addressed to D.A.H. (dhinds@perlegen.com).

Two recent genome-wide association studies have described associations of SNP variants in *PNPLA3* with nonalcoholic fatty liver and plasma liver enzyme levels in population based cohorts. We investigated the contributions of these variants to clinical outcomes in Mestizo subjects with a history of excessive alcohol consumption. We show that non-synonymous variant rs738409[G] (I148M) in *PNPLA3* is strongly associated with alcoholic liver disease and progression to alcoholic cirrhosis (unadjusted OR = 2.25, $P = 1.7\times10^{-10}$; ancestry-adjusted OR = 1.79, $P = 1.9\times10^{-5}$).

Liver cirrhosis is the twelfth leading cause of death in the United States, with a total of 28,175 deaths reported in 2005[1]. Nearly half of these were classified as alcohol-related, yet this cause may be substantially under-reported[2]. Hispanics are disproportionately impacted by chronic liver disease[3]. In 2007, cirrhosis was the fourth leading cause of death in Mexico, and the second leading cause in adults age 15 to 64 (http://sinais.salud.gob.mx/mortalidad/). Only 10% to 15% of alcoholics develop cirrhosis, and while patterns of alcohol consumption are clearly important, they do not appear to fully account for the ethnic differences in incidence rates[4].

Recently, Romeo *et al.*[5] carried out a genome-wide association study of non-synonymous sequence variations in a population comprising Hispanics, African Americans and European Americans to identify genetic variants contributing to nonalcoholic fatty liver disease (NAFLD). They found strong evidence of association of an allele in *PNPLA3* (rs738409[G], I148M) with increased hepatic fat levels ($P = 5.9\times10^{-10}$) and the association remained highly significant after adjustment for body mass index, diabetes status, ethanol use and genetic ancestry. The study also revealed a significant elevation in serum concentrations of alanine aminotransferase (ALT) in association with the rs738409[G] allele in Hispanics ($P = 1.3\times10^{-5}$). Resequencing the region revealed another allele (rs6006460[T], S453I) that was independently associated with lower

hepatic fat content in African Americans, the group at lowest risk of NAFLD. An independent genome-wide study[6] aiming to find genes influencing plasma levels of liver enzymes also showed evidence that variants in *PNPLA3* were associated with plasma levels of alanine-aminotransferase (ALT). This study gave additional support to the hypothesis that rs738409[G] may confer increased susceptibility to hepatic injury.

We attempted to follow up on these findings in Mestizo individuals from Mexico City with a history of alcohol abuse. In addition to the reported non-synonymous variants rs738409 and rs6006460, we assayed 15 common tagging SNPs from the *PNPLA3* region, 291 SNPs for assessing global ancestry[7], 16 ancestry-informative markers (AIMs) flanking the *PNPLA3* region for assessing local ancestry[8], and 7 SNPs previously reported to be associated with cirrhosis in hepatitis C patients[9]. We successfully genotyped 305 individuals with apparently normal liver function, 434 with intermediate alcoholic liver disease (ALD), and 482 with alcoholic cirrhosis (Supplementary Methods online).

The clinical characteristics of our samples are shown in Table 1. The healthy, alcoholic liver disease and cirrhosis diagnosis groups had significant differences in mean age, mean alcohol intake and duration, and mean global and local loadings from a principal component analysis (PCA) of the genotyping data. Both local and global mean individual Native American ancestry were significantly higher among the cirrhosis individuals than the healthy individuals ($P < 2.2 \times 10^{-16}$), consistent with the higher prevalence of cirrhosis in Hispanics compared to individuals of European or African ancestry[3].

We tested each SNP for association with cirrhosis or ALD using likelihood ratio tests from logistic regression, adjusted for age, alcohol intake and duration, an interaction between age and duration, and global genetic ancestry estimated using PCA. The association analysis was

carried out for three pair-wise combinations of diagnosis results: cirrhosis versus healthy, cirrhosis versus ALD, and ALD versus healthy. We also performed tests controlling for local ancestry along the 9.4 Mb region flanking rs738409 estimated from PCA using the 16 AIMs. SNP rs738409 is strongly associated with alcoholic cirrhosis (Table 2).

In tests of cirrhosis versus healthy status, rs738409 showed strong association before and after controlling for global ancestry, as well as after controlling for local ancestry. Test results for cirrhosis versus ALD, and for ALD versus healthy status, suggest that rs738409 has an intermediate effect on the ALD phenotype. We used the Akaike information criterion (AIC) to compare four genetic models (2-df general model, and 1-df additive, dominant, and recessive risk allele models) in logistic regression adjusting for covariates and the global individual ancestry. The most parsimonious model was an additive model for rs738409 [G] (Supplementary Table 2 online). Logistic regression analysis suggested that the rs738409 sequence variation accounts for 49% of the observed ancestry-related difference in cirrhosis susceptibility (See Supplementary Methods). Further tests showed no interactions of rs738409 with other covariates, including age, alcohol intake and duration.

Association test results for the tagging SNPs were generally consistent with their extent of linkage disequilibrium (LD) with rs738409 (Supplementary Figure 1 and Table 1 online). SNP rs738408 is three base pairs away from rs738409 and is in nearly complete linkage disequilibrium (LD) ($r^2 = 0.99$). Association tests for rs738408 gave nearly identical results as rs738409. When the SNP rs738409 was treated as causal by including its genotypes as a covariate in the regression model, the additional associations in the *PNPLA3* region were eliminated. All common haplotypes containing the rs738409[G] allele were more common in cirrhosis cases than in the healthy group (Supplementary Table 3 online). The rare variant

rs6006460[T] reported by Romeo *et al.*[5] is also rare in our Mestizo population (MAF = 0.002) so we had no power to detect an effect of this variant. The minor T allele of rs6006460 was observed in both the cirrhosis and healthy groups. We did not observe significant associations with any of the markers previously reported to be associated with cirrhosis in hepatitis C patients[9].

We further tested for an association of rs738409 with prognosis in cirrhosis patients, as indicated by Child-Pugh class[10]. We coded Child-Pugh classes as numeric scores from 1 to 3 in order of severity, and fit by linear regression with the same covariates used in the binary outcome models and an additive genotype term. The high-risk G allele showed a suggestive association with increasing severity (Wald test: one-sided $P = 0.05$). The frequencies of rs738409[G] in patients scored as Child-Pugh class A, B, and C were 0.70, 0.75 and 0.77, respectively.

While the biochemical function of *PNPLA3* is unclear, it is primarily expressed in adipocytes and may have a role in energy homeostasis[11]. Our results suggest that individual differences in lipid metabolism can predispose to both alcoholic and non-alcoholic liver injury. This further supports a central role for altered lipid metabolism in liver pathogenesis[12]. Variation in *PNPLA3* has also been associated with obesity and insulin resistance[13]. We did not have access to metabolic phenotypes in our population, but given these results, it may be useful to investigate interactions with body mass, lipid profiles, and insulin sensitivity.

Our study extends the previously reported associations of rs738409 with subclinical nonalcoholic liver disease to clinically relevant endpoints of alcoholic liver disease. This single variant accounts for a substantial share of the increased risk of cirrhosis associated with Hispanic ancestry. Hispanics with hepatitis C are also at substantially elevated risk for hepatic injury

compared to other ethnicities[14], and it will be important to determine if rs738409 is associated in that context as well. The effect size of rs738409 is large for an association with complex disease in humans, and may be the largest known genetic modifier for a disease that is a major cause of preventable death. For these reasons, this variant may be an attractive target for genetic screening to identify individuals at high risk for liver disease for more aggressive interventions.

Author contributions

D.G.B. participated in the conception and design of the study. D.K. provided the samples used in the study. R.P.S. directed the genotyping experiments. D.A.H. and C.T. analyzed the data and drafted the manuscript. All authors contributed to the final manuscript.

Acknowledgments

We thank the individual participants in the study who made this work possible. We thank D. Cox, N. Patil, N. Shen, and C. Hacker for their work to obtain and characterize the samples. We also thank E. Beasley and M. Mealiffe for guidance, and E. Beilharz, T. Genschoreck, M. Kelleher, M. Kennemer, A. Ollmann, K. Pant, and S. Pitts for technical assistance.

References

1. Yoon, Y.H. & Yi, H.Y. Surveillance Report #83: Liver Cirrhosis Mortality in the United States, 1970-2005. National Institute on Alcohol Abuse and Alcoholism (2008).
2. Baker, A., & Rooney, C. *Health Stat. Q.* **17**, 5-14 (2003).
3. Flores, Y.N. *et al. Am. J. Gastroenterol.* **103**, 2231-2238 (2008).
4. Mann, R.E., Smart, R.G., & Govoni, R. *J. Natl. Inst. Alcohol Abuse Alcohol.* **27**, 209-219 (2003).
5. Romeo, S. *et al. Nat. Genet.* **40**, 1461-1465 (2008).
6. Yuan, X. *et al. Am. J. Hum. Genet.* **83**, 520-528 (2008).
7. Hinds, D.A. *et al. Am. J. Hum. Genet.* **74**, 317-325 (2004).
8. Tian, C. *et al. Am. J. Hum. Genet.* **80**, 1014-1023 (2007).
9. Huang, H. *et al. Hepatology* **46**, 297-306 (2007).
10. Pugh, R., Murray-lyon, I., & Dawson, J. *Br. J. Surg.* **60**, 646-649 (1973).
11. Baulande, S., Lasnier, F., Lucas, M., & Pairault, J. *J. Biol. Chem.* **276**, 33336-33344 (2001).
12. Sozio, M. & Crabb, D.W. *Am. J. Physiol. Endocrinol. Metab.* **295**, E10-E16 (2008).
13. Johansson, L.E., Lindblad, U., Larsson, C.A., Råstam, L., & Ridderstråle, M. *Eur. J. Endocrinol.* **159**, 577-583 (2008).
14. Hoyumpa, A., *et al. Hepatology* **42(suppl 1)**, A442 (2005).

Table 1. Clinical characteristics of the genotyped subjects.

Characteristic	Healthy	ALD	Cirrhosis	*P* value
Gender: n (%)				0.92
F	48 (0.16)	67 (0.15)	71 (0.15)	
M	257 (0.84)	367 (0.85)	411 (0.85)	
Intake, gm/week: n (%)				9.4×10^{-11}
200-500	186 (0.61)	162 (0.37)	179 (0.37)	
500-750	20 (0.07)	46 (0.11)	48 (0.10)	
>750	99 (0.32)	226 (0.52)	255 (0.53)	
Duration, years: mean (sd)	17.5 (9.62)	19.5 (9.89)	26.1 (11.42)	$<2.2\times10^{-16}$
Age, years: mean (sd)	39 (12.7)	41 (12.4)	52 (11.6)	$<2.2\times10^{-16}$
Global PC1 loading: mean (sd)	-0.0098 (0.027)	-0.00004 (0.024)	0.0058 (0.026)	$<2.2\times10^{-16}$
Local PC1 loading: mean (sd)	-0.0094 (0.029)	0.00106 (0.026)	0.0049 (0.027)	5.3×10^{-12}

P values are from F tests for continuous outcomes, and Pearson χ^2 tests for categorical outcomes.

Table 2. Association test results for rs738409.

	Genotype Count						Allele Frequency			
	Control			Case			Control	Case	Allelic Test	
	CC	CG	GG	CC	CG	GG	G	G	OR (95% CI)	*P*
Cirrhosis vs Healthy	83	198	111	59	264	371	0.54	0.72	2.28 (1.90-2.74)	7.6×10^{-19}
Cirrhosis vs ALD	91	266	305	59	264	371	0.66	0.72	1.35 (1.14-1.59)	4.2×10^{-4}
ALD vs Healthy	83	198	111	91	266	305	0.54	0.66	1.69 (1.41-2.03)	0.0012

	Logistic Regression Likelihood Ratio Tests					
	No Ancestry Correction		Global Correction		Global and Local Correction	
	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Cirrhosis vs Healthy	2.25 (1.74-2.90)	1.7×10^{-10}	1.79 (1.37-2.35)	1.9×10^{-5}	1.81 (1.36-2.41)	4.7×10^{-5}
Cirrhosis vs ALD	1.43 (1.15-1.78)	0.0010	1.33 (1.06-1.66)	0.014	1.45 (1.13-1.84)	0.0028
ALD vs Healthy	1.45 (1.16-1.80)	8.4×10^{-4}	1.26 (1.00-1.58)	0.051	1.18 (0.92-1.51)	0.19

Genotypes describe the forward strand on NCBI Build 36, and odds ratios are per [G] allele.

Variation in *PNPLA3* is associated with outcomes in alcoholic liver disease

Chao Tian, Renee P. Stokowski, David Kershenobich, Dennis G. Ballinger, David A. Hinds

Supplementary Methods

Samples

Subjects were recruited from clinics in Mexico City based on a history of alcohol abuse. Patients with liver disease were enrolled from the liver, gastroenterology, and internal medicine clinics at INCMNSZ. Alcoholic controls were enrolled from the Institute of Nutrition at INCMNSZ, as well as the Instituto Mexicano de Psiquiatria (IMP) and the Centro de Ayuda al Alcoholico y sus Familiares (CAAF). We excluded individuals with other than self-reported Mestizo ancestry. Patients with viral hepatitis, or systemic chronic disease (cardiac insufficiency, HIV or autoimmune disease, chronic renal insufficiency, or neoplastic disease) were excluded from the study. Subjects were assessed for overt liver disease using a combination of clinical and biochemical criteria:

- Clinical criteria: jaundice, spider angiomas, palmar erythema, abnormal collateral circulation, ascitis, hepatic encephalopathy, esophageal varices, portal hypertension
- Biochemical criteria: abnormal aminotransferases, glutamil transpeptidase, alkaline phosphatase, decreased serum albumin, increased serum globulin, decreased prothrombin

Subjects were classified as “healthy” if they had no indicators of liver disease. For patients with liver disease, ultrasound or CT screening was used to corroborate diagnoses of cirrhosis. Subjects were also assessed for fatty liver and alcoholic hepatitis. Patients with liver disease that did not meet the criteria for cirrhosis were classified as “alcoholic liver disease”. Typical alcohol

consumption and duration (prior to diagnosis for cirrhosis patients) was determined by interview. We excluded individuals reporting fewer than 5 years or less than 200 gm/week of alcohol consumption.

Genotyping

We genotyped samples on microarrays across 312 autosomal SNPs uniformly distributed across the genome for assessment of global population structure. We eliminated microarray scans with call rate < 90% and SNPs with call rate < 95%. The microarray data was collected in several batches processed at different times. We clustered the combined data together and tested all SNPs for batch effects by ANOVA. We identified 2 SNPs with strong batch effects ($P < 10^{-70}$ and $r^2 > 0.2$), and an additional 7 SNPs with smaller effects ($P < 0.001$ and $r^2 > 0.01$). We manually inspected cluster plots for these SNPs to confirm that these SNPs did show clustering artifacts, and that SNPs with $P > 0.001$ appeared to be unaffected. We also removed SNPs with Hardy Weinberg $P < 10^{-7}$, based again on manual inspection of cluster diagrams. Across 291 SNPs passing these filters, the mean call rate was 99.9%.

Separately, the samples were genotyped across 61 SNPs on the MassArray platform by Sequenom. These included the two non-synonymous variants rs738409 and rs6006460 reported in Romeo *et al.*; 17 common tagging SNPs from the region surrounding the *PNPLA3* gene; 16 ancestry informative markers (AIMs) flanking *PNPLA3* for assessing local admixture; 7 SNPs previously reported to be associated with cirrhosis in patients with chronic hepatitis C; and 18 quality control (QC) SNPs selected from the microarray panel for verification of sample identities. We also included rs738408, a synonymous SNP 3 base pairs away from and in perfect LD with rs738409 in the HapMap CEU panel. Where possible, we designed redundant assays in opposite orientations, and selected the assay with highest call rate for each SNP. We excluded samples with call rate < 80%, SNPs with call rate < 90%, and SNPs with Hardy Weinberg $P < 10^{-7}$. We used more generous call rate thresholds for the Sequenom data because missing data rates were more broadly distributed than in the microarray data, and these SNPs were considered

to be of higher value. There were 4 SNPs (2 tag SNPs, 2 QC SNPs) for which no assays passed the quality filters. The remaining 57 SNPs had an average call rate of 98.7%.

We identified 36 pairs of samples that had nearly identical microarray genotype patterns. These cryptic duplicates tended to have correlated ages ($r^2 = 0.3$, $P = 0.0002$) and diagnoses (Fisher's exact test: $P = 0.00003$), suggesting that in at least some cases the same individual had been recruited twice. An additional 21 samples had low concordance across the 16 SNPs successfully genotyped on both platforms. These two groups of samples were excluded from analyses. Across the remaining samples, the 16 shared SNPs had an average concordance of 99.9% across platforms.

Statistical Analysis

The significance of the differences in mean sample characteristics such as gender, age, alcoholic intake and duration among diagnostic groups were tested using ANOVA. Tests for Hardy-Weinberg equilibrium were performed for each SNP in each diagnosis group using a likelihood ratio test (LRT). All of the tested SNPs were in Hardy-Weinberg equilibrium ($P > 0.001$) in each group.

Association Tests

Allelic odds ratios (OR) were estimated from 2×2 contingency tables, and approximate 95% confidence intervals (CI) for odds ratios were estimated using Woolf's method[1]. The allelic association test is the 1 degree of freedom Pearson Chi-square test of the allele frequency differences between case and control subjects. A stepwise logistic regression using Akaike information criterion (AIC) was also used to define the confounders with disease risk. Age, alcohol intake and duration, and an interaction between age and duration were identified as significant confounders based on AIC. ORs for cirrhosis risk under an additive genetic model were estimated using logistic regression with adjustment for these confounders. The model for our primary single SNP association tests can be written as:

$$\text{Logit}(\pi) = \beta_0 + \beta_1 * \text{age} + \beta_2 * \text{duration} + \beta_3 * \text{intake} + \beta_4 * (\text{age x duration}) + \beta_5 * \text{genotype}$$

where π is the likelihood of membership in the case group. The LRT compared models with and without the genotype covariate. Detailed results are shown in Supplementary Table 1.

Estimating Genetic Ancestry

The genetic diversity due to admixture within the Mestizo population is a known confounder for our association tests. To control for spurious association due to differences in ancestry between diagnosis groups, we inferred individual global and local ancestry using Principal Components Analysis (PCA). Global ancestry was determined for each individual using 291 SNPs distributed across the genome. Local ancestry was determined for each individual using 16 AIMs for Amerindian and European genetic ancestry flanking the *PNPLA3* locus (Supplementary Table 4). We incorporated data from additional individuals with self-reported Otomi Indian and European ancestry to facilitate interpretation of the PCA results. Individuals with more Native American ancestry have larger loadings along the first principal component (Supplementary Figure 2). We also applied PCA to just the Mestizo individuals, and found that PC1 had a correlation of 0.9995 with corresponding results from PCA with the additional samples. PC1 explained substantially more genetic variance than the higher order components (Supplementary Figure 3). ANOVA tests of the loadings across diagnosis groups show that only PC1 is substantially associated with outcomes, for both local and global PCA results. Both local and global ancestry estimation were also confirmed by the Bayesian Markov Chain-Monte Carlo (MCMC) method implemented in the program STRUCTURE[2]. STRUCTURE 2.1 was run under the admixture model and k=3 for global ancestry estimation, and linkage model and k=2 for local ancestry estimation. We ran the MCMC method with burn-in length of 20,000 for 20,000 repetitions. The admixture proportions estimated for the most representative ancestry group had a correlation of 0.998 with PC1 of the corresponding PCA.

It is broadly known that the Mestizo population has an admixture of Native American, European and African ancestry. We used STRUCTURE to obtain ancestry proportions for our

combined Otomi, European, and Mestizo data, together with data from the HapMap YRI panel for the same 291 SNPs (Supplementary Figure 4). The results shows that our Mestizo samples have about 36.5% (sd 0.214) European, 61.6% (0.22) Native American and 1.9% (0.033) African ancestry on average, consistent with a previous report[3]. The mean individual African ancestry estimated using STRUCTURE software is not significantly different among the diagnosis groups and is small (< 2%). Thus, we expect to see little confounding in our association studies due to African admixture. Association tests controlling for the STRUCTURE results gave nearly identical results as controlling for the first PC from global PCA. The first PC has a correlation of 0.99 with the estimated Native American or European proportion from STRUCTURE.

Structured Association Tests

We adjusted for differences in individual global ancestry in association tests by including the first PC from global PCA as a covariate in the logistic regression. These tests indicate the strength of association attributable to a SNP that is independent of information that SNP might indirectly provide about overall ancestry across the genome. To verify that associations could not be explained by confounding with local admixture, we also tested SNPs in the *PNPLA3* region with adjustment for local ancestry by additionally including the first PC from local PCA as a covariate. Our models for structured association tests can be written as:

$$\text{Logit}(\pi) = \beta_0+\beta_1*\text{age}+\beta_2*\text{duration}+\beta_3*\text{intake}+\beta_4*(\text{age}*\text{duration})+\beta_5*\text{GPC1}+\beta_6*\text{genotype}$$

$$\text{Logit}(\pi) = \beta_0+\beta_1*\text{age}+\beta_2*\text{duration}+\beta_3*\text{intake}+\beta_4*(\text{age}*\text{duration})+\beta_5*\text{GPC1}+\beta_6*\text{LPC1}+\beta_7*\text{genotype}$$

where GPC1 is the first PC from global PCA and LPC1 is the first PC from local PCA. The LRT compared models with and without the genotype covariate. Detailed results for each SNP are shown in Supplementary Table 1.

To study the genetic risk models for rs738409, we used the AIC to compare four possible genetic models: a 2-degree-of-freedom general model, and 1-degree-of-freedom additive, dominant, and recessive models. The three possible genotypes [AA,AB,BB] of tested SNPs were coded as a three-level factor for the general model, [0,1,2] for the additive model, [0,0,1] for the

dominant model, and [0,1,1] for the recessive model. The most parsimonious model was defined as the one with the smallest AIC.

We also assessed the ability of rs738409 to explain ancestry differences in cirrhosis susceptibility. Since the rs738409 sequence variation is substantially correlated with global ancestry ($r = 0.34$), we calculated ancestry-adjusted genotypes in which redundancies between rs738409 and ancestry were removed, and fit four logistic regression models:

(1) $\text{Logit}(\pi) = \beta_0+\beta_1*\text{age}+\beta_2*\text{duration}+\beta_3*\text{intake}+\beta_4*(\text{age}*\text{duration})+\beta_5*\text{GPC1}+\beta_6*\text{rs738409}$,

(2) $\text{Logit}(\pi) = \beta_0+\beta_1*\text{age}+\beta_2*\text{duration}+\beta_3*\text{intake}+\beta_4*(\text{age}*\text{duration})+\beta_5*\text{rs738409}$

(3) $\text{Logit}(\pi) = \beta_0+\beta_1*\text{age}+\beta_2*\text{duration}+\beta_3*\text{intake}+\beta_4*(\text{age}*\text{duration})+\beta_5*\text{GPC1}+\beta_6*\text{rs738409\_adj}$,

(4) $\text{Logit}(\pi) = \beta_0+\beta_1*\text{age}+\beta_2*\text{duration}+\beta_3*\text{intake}+\beta_4*(\text{age}*\text{duration})+\beta_5*\text{rs738409\_adj}$

Here, rs738409_adj represents residuals from linear regression of rs738409 genotypes against global PC1. The proportion of residual deviance explained by global PC1 estimated from models (3) and (4) is 6.43%, while the proportion explained by global PC1 estimated from models (1) and (2) is 3.28%. Thus, rs738409 accounted for 49% of the observed ancestry-related difference in cirrhosis susceptibility.

Haplotype Analyses

We used MACH[4] version 1.1 to phase genotype data for 18 SNPs in the *PNPLA3* region. We performed association tests for each 18-SNP haplotype with observed MAF > 0.01, versus all other haplotypes, both using Pearson χ^2 tests on 2×2 tables, and logistic regression as in the single SNP analyses (Supplementary Table 3).

References

1. Woolf, B. *Ann. Hum. Genet.* **19**, 251-253 (1955).
2. Falush, D., Stephens, M., & Pritchard, J.K. *Genetics* **164**, 1567-1587 (2003).
3. Wang, S *et al. PLOS Genet.* **4(3)**, e1000037. doi:10.1371/journal.pgen.1000037 (2008).
4. Li, Y. & Abecasis, G.R. *Am. J. Hum. Genet.* **S79**, 2290 (2006).
5. Barrett, J.C., Fry, B., Maller, J., & Daly, M.J. *Bioinformatics* **21**, 263-265 (2005).

Supplementary Table 1a. Cirrhosis versus Healthy Association Tests.

					Alleles		Alt Frequency		Allelic Test	
rsID	Chromosome	Position	Category	LD	Ref	Alt	Control	Case	P	OR (95% CI)
rs4290029	1	222467263	Huang		C	G	0.69	0.68	8.0E-01	0.97 (0.80–1.18)
rs17740066	3	122582973	Huang		G	A	0.17	0.23	1.7E-03	1.47 (1.16–1.86)
rs62522600	8	103910885	Huang		G	A	0.02	0.02	1.7E-01	0.61 (0.33–1.15)
rs4986791	9	119515423	Huang		C	T	0.02	0.02	6.8E-01	1.23 (0.62–2.45)
rs886277	11	2396343	Huang		T	C	0.52	0.57	1.5E-02	1.25 (1.05–1.49)
rs2878771	12	48638660	Huang		G	C	0.12	0.08	2.8E-03	0.64 (0.48–0.85)
rs2290351	15	88175785	Huang		G	A	0.37	0.42	1.4E-02	1.26 (1.05–1.51)
rs9614194	22	42636398	LDtag	0.027	G	A	0.04	0.01	2.0E-04	0.34 (0.20–0.61)
rs5764023	22	42636551	LDtag	0.031	C	T	0.20	0.25	9.2E-03	1.34 (1.08–1.67)
rs4823168	22	42638868	LDtag	0.000	C	T	0.27	0.24	1.3E-01	0.85 (0.70–1.04)
rs929090	22	42645182	LDtag	0.010	A	G	0.55	0.56	7.0E-01	1.04 (0.86–1.26)
rs4823104	22	42652482	LDtag	0.021	A	G	0.05	0.03	1.5E-01	0.70 (0.45–1.09)
rs2076213	22	42654255	LDtag	0.070	T	G	0.19	0.26	2.9E-04	1.50 (1.21–1.86)
rs2076212	22	42654303	LDtag	0.037	G	T	0.10	0.07	7.7E-03	0.64 (0.47–0.88)
rs738407	22	42655288	LDtag	0.531	T	C	0.60	0.75	1.6E-12	2.03 (1.67–2.47)
rs139051	22	42656009	LDtag	0.473	A	G	0.38	0.23	1.3E-13	0.49 (0.40–0.59)
rs738409	22	42656060	Romeo	1.000	C	G	0.54	0.72	7.7E-19	2.28 (1.90–2.74)
rs738408	22	42656063	Romeo	0.988	C	T	0.54	0.72	2.3E-17	2.21 (1.84–2.65)
rs1883350	22	42659376	LDtag	0.776	T	C	0.56	0.73	4.7E-15	2.11 (1.75–2.54)
rs4823173	22	42660063	LDtag	0.833	G	A	0.49	0.69	5.7E-19	2.28 (1.90–2.73)
rs2076208	22	42662393	LDtag	0.273	G	C	0.84	0.88	3.2E-02	1.33 (1.03–1.71)
rs2294916	22	42672255	LDtag	0.791	T	G	0.49	0.67	1.4E-16	2.14 (1.79–2.57)
rs2294918	22	42673449	LDtag	0.352	A	G	0.76	0.87	2.0E-09	2.02 (1.60–2.54)
rs6006460	22	42673507	Romeo	0.008	G	T	0.00	0.00	5.1E-01	0.37 (0.06–2.24)
rs2294919	22	42673658	LDtag	0.279	C	T	0.17	0.13	9.5E-03	0.72 (0.56–0.92)

	Logistic Regression Likelihood Ratio Tests							
	No Ancestry Correction		Global Correction		Global & Local Correction		Condition on rs738409	
rsID	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)
rs4290029	4.5E-01	1.11 (0.85–1.45)	7.0E-02	1.29 (0.98–1.71)				
rs17740066	8.8E-03	1.49 (1.10–2.02)	6.3E-02	1.35 (0.98–1.86)				
rs62522600	5.7E-02	0.43 (0.18–1.03)	3.3E-01	0.64 (0.26–1.57)				
rs4986791	4.2E-01	0.70 (0.30–1.65)	9.0E-01	0.94 (0.38–2.32)				
rs886277	7.2E-02	1.25 (0.98–1.59)	3.5E-01	1.13 (0.88–1.45)				
rs2878771	1.3E-03	0.52 (0.35–0.78)	3.5E-02	0.64 (0.43–0.97)				
rs2290351	7.1E-02	1.25 (0.98–1.59)	7.1E-01	1.05 (0.81–1.35)				
rs9614194	6.6E-04	0.29 (0.14–0.60)	5.9E-03	0.36 (0.17–0.75)	1.7E-02	0.41 (0.20–0.86)	7.0E-02	0.49 (0.23–1.07)
rs5764023	2.9E-01	1.17 (0.87–1.57)	8.8E-01	1.02 (0.75–1.39)	9.4E-01	1.01 (0.75–1.37)	5.6E-01	0.91 (0.66–1.25)
rs4823168	5.7E-01	0.92 (0.70–1.21)	9.0E-01	0.98 (0.74–1.30)	9.0E-01	1.02 (0.77–1.35)	9.6E-01	1.01 (0.76–1.34)
rs929090	6.8E-01	1.05 (0.82–1.35)	8.3E-01	0.97 (0.75–1.26)	7.6E-01	0.96 (0.74–1.24)	6.1E-01	0.93 (0.72–1.22)
rs4823104	1.9E-01	0.67 (0.38–1.20)	9.3E-01	0.97 (0.54–1.76)	8.8E-01	0.95 (0.52–1.75)	9.6E-01	1.02 (0.55–1.88)
rs2076213	8.4E-02	1.28 (0.96–1.71)	6.9E-01	1.06 (0.79–1.43)	8.6E-01	1.03 (0.76–1.39)	5.9E-01	0.92 (0.67–1.25)
rs2076212	1.2E-02	0.59 (0.40–0.89)	4.6E-02	0.65 (0.43–0.99)	3.6E-02	0.64 (0.42–0.97)	2.3E-01	0.77 (0.50–1.18)
rs738407	1.8E-06	1.87 (1.44–2.42)	1.1E-02	1.44 (1.09–1.90)	3.2E-02	1.39 (1.03–1.87)	5.3E-01	0.88 (0.60–1.30)
rs139051	2.0E-06	0.53 (0.41–0.69)	8.6E-03	0.69 (0.52–0.91)	4.1E-02	0.73 (0.53–0.99)	9.3E-01	1.02 (0.70–1.47)
rs738409	1.7E-10	2.25 (1.74–2.90)	1.9E-05	1.79 (1.37–2.35)	4.7E-05	1.81 (1.36–2.41)		
rs738408	3.0E-10	2.25 (1.74–2.91)	2.8E-05	1.79 (1.36–2.35)	5.5E-05	1.81 (1.35–2.42)	9.9E-01	1.01 (0.11–9.15)
rs1883350	6.4E-07	1.83 (1.44–2.32)	4.2E-03	1.45 (1.13–1.88)	1.2E-02	1.42 (1.08–1.86)	2.0E-01	0.71 (0.42–1.20)
rs4823173	9.9E-10	2.10 (1.64–2.67)	1.8E-04	1.65 (1.27–2.14)	4.9E-04	1.67 (1.25–2.22)	1.0E+00	1.00 (0.55–1.82)
rs2076208	1.2E-01	1.31 (0.93–1.84)	4.0E-01	1.16 (0.82–1.65)	2.8E-01	1.21 (0.86–1.72)	1.4E-01	0.73 (0.49–1.11)
rs2294916	8.2E-08	1.91 (1.50–2.43)	2.0E-03	1.50 (1.16–1.95)	6.9E-03	1.48 (1.11–1.96)	3.5E-01	0.78 (0.46–1.32)
rs2294918	1.4E-04	1.87 (1.35–2.58)	2.9E-02	1.46 (1.04–2.04)	1.1E-01	1.34 (0.94–1.91)	7.9E-01	0.95 (0.63–1.43)
rs6006460	5.8E-01	0.49 (0.04–5.52)	6.8E-01	0.59 (0.05–6.72)	9.4E-01	1.10 (0.09–13.3)	9.1E-01	0.86 (0.07–10.2)
rs2294919	1.9E-02	0.67 (0.48–0.94)	1.3E-01	0.77 (0.54–1.08)	6.9E-02	0.73 (0.52–1.02)	4.0E-01	1.19 (0.80–1.77)

Chromosomal positions and alleles are given for the forward strand of NCBI Build 36. 'LD' indicates r^2 with rs738409. Odds ratios are specified per alternate allele.

Supplementary Table 1b. Cirrhosis versus ALD Association Tests.

					Alleles		Alt Frequency		Allelic Test	
rsID	Chromosome	Position	Category	LD	Ref	Alt	Control	Case	P	OR (95% CI)
rs4290029	1	222467263	Huang		C	G	0.69	0.68	3.6E-01	1.08 (0.92–1.28)
rs17740066	3	122582973	Huang		G	A	0.17	0.23	2.9E-02	1.25 (1.03–1.53)
rs62522600	8	103910885	Huang		G	A	0.02	0.02	6.3E-01	0.83 (0.46–1.49)
rs4986791	9	119515423	Huang		C	T	0.02	0.02	8.5E-01	0.91 (0.53–1.57)
rs886277	11	2396343	Huang		T	C	0.52	0.57	4.9E-01	0.94 (0.81–1.10)
rs2878771	12	48638660	Huang		G	C	0.12	0.08	3.1E-01	0.86 (0.66–1.13)
rs2290351	15	88175785	Huang		G	A	0.37	0.42	8.7E-01	0.98 (0.85–1.15)
rs9614194	22	42636398	LDtag	0.027	G	A	0.04	0.01	7.5E-01	0.86 (0.47–1.59)
rs5764023	22	42636551	LDtag	0.031	C	T	0.20	0.25	3.2E-01	1.10 (0.92–1.32)
rs4823168	22	42638868	LDtag	0.000	C	T	0.27	0.24	5.3E-01	1.06 (0.89–1.27)
rs929090	22	42645182	LDtag	0.010	A	G	0.55	0.56	5.8E-01	1.05 (0.89–1.23)
rs4823104	22	42652482	LDtag	0.021	A	G	0.05	0.03	8.2E-01	0.93 (0.62–1.40)
rs2076213	22	42654255	LDtag	0.070	T	G	0.19	0.26	2.9E-02	1.22 (1.02–1.46)
rs2076212	22	42654303	LDtag	0.037	G	T	0.10	0.07	7.8E-01	0.95 (0.70–1.28)
rs738407	22	42655288	LDtag	0.531	T	C	0.60	0.75	4.0E-02	1.21 (1.01–1.45)
rs139051	22	42656009	LDtag	0.473	A	G	0.38	0.23	1.8E-01	0.88 (0.74–1.05)
rs738409	22	42656060	Romeo	1.000	C	G	0.54	0.72	4.2E-04	1.35 (1.14–1.59)
rs738408	22	42656063	Romeo	0.988	C	T	0.54	0.72	1.3E-03	1.31 (1.11–1.55)
rs1883350	22	42659376	LDtag	0.776	T	C	0.56	0.73	3.3E-03	1.29 (1.09–1.52)
rs4823173	22	42660063	LDtag	0.833	G	A	0.49	0.69	2.8E-03	1.28 (1.09–1.51)
rs2076208	22	42662393	LDtag	0.273	G	C	0.84	0.88	1.5E-03	1.43 (1.15–1.77)
rs2294916	22	42672255	LDtag	0.791	T	G	0.49	0.67	6.5E-03	1.25 (1.07–1.47)
rs2294918	22	42673449	LDtag	0.352	A	G	0.76	0.87	7.1E-01	1.05 (0.84–1.31)
rs6006460	22	42673507	Romeo	0.008	G	T	0.00	0.00	5.8E-02	0.21 (0.05–0.97)
rs2294919	22	42673658	LDtag	0.279	C	T	0.17	0.13	1.1E-03	0.70 (0.57–0.86)

	Logistic Regression Likelihood Ratio Tests							
	No Ancestry Correction		Global Correction		Global & Local Correction		Condition on rs738409	
rsID	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)
rs4290029	9.7E-01	1.00 (0.80–1.24)	6.9E-01	1.05 (0.84–1.31)				
rs17740066	7.7E-02	1.26 (0.97–1.64)	1.5E-01	1.21 (0.93–1.57)				
rs62522600	8.1E-01	0.90 (0.39–2.10)	9.3E-01	1.04 (0.44–2.44)				
rs4986791	5.5E-01	0.80 (0.39–1.64)	6.8E-01	0.86 (0.42–1.76)				
rs886277	3.5E-01	0.91 (0.74–1.11)	2.0E-01	0.87 (0.71–1.07)				
rs2878771	4.7E-02	0.70 (0.49–1.00)	1.2E-01	0.75 (0.52–1.08)				
rs2290351	5.8E-01	1.06 (0.86–1.30)	9.6E-01	0.99 (0.80–1.23)				
rs9614194	6.9E-01	0.86 (0.41–1.82)	9.5E-01	0.98 (0.45–2.10)	9.1E-01	0.96 (0.44–2.06)	8.2E-01	1.09 (0.51–2.35)
rs5764023	2.4E-01	1.16 (0.91–1.48)	3.8E-01	1.12 (0.87–1.43)	3.0E-01	1.14 (0.89–1.46)	6.1E-01	1.07 (0.83–1.37)
rs4823168	5.6E-01	1.07 (0.84–1.36)	4.1E-01	1.11 (0.87–1.41)	4.5E-01	1.10 (0.86–1.40)	4.2E-01	1.10 (0.87–1.41)
rs929090	4.1E-01	1.09 (0.89–1.34)	5.2E-01	1.07 (0.87–1.32)	4.6E-01	1.08 (0.88–1.33)	5.8E-01	1.06 (0.86–1.31)
rs4823104	5.0E-01	0.84 (0.51–1.39)	9.4E-01	0.98 (0.59–1.64)	7.7E-01	0.92 (0.54–1.57)	9.0E-01	1.04 (0.62–1.74)
rs2076213	7.3E-02	1.25 (0.98–1.60)	2.4E-01	1.17 (0.90–1.50)	1.3E-01	1.22 (0.94–1.57)	5.0E-01	1.09 (0.84–1.42)
rs2076212	7.5E-01	1.07 (0.72–1.59)	5.6E-01	1.13 (0.75–1.68)	6.3E-01	1.10 (0.74–1.65)	3.0E-01	1.24 (0.82–1.87)
rs738407	3.6E-02	1.28 (1.02–1.63)	2.1E-01	1.17 (0.91–1.50)	8.1E-02	1.28 (0.97–1.68)	5.2E-01	0.89 (0.63–1.26)
rs139051	2.8E-02	0.77 (0.61–0.97)	2.2E-01	0.85 (0.67–1.10)	8.5E-02	0.79 (0.60–1.03)	6.8E-01	1.07 (0.77–1.48)
rs738409	1.0E-03	1.43 (1.15–1.78)	1.4E-02	1.33 (1.06–1.66)	2.8E-03	1.45 (1.13–1.84)		
rs738408	1.8E-03	1.41 (1.13–1.75)	2.1E-02	1.30 (1.04–1.63)	4.8E-03	1.42 (1.11–1.81)	3.1E-01	0.36 (0.04–3.18)
rs1883350	1.5E-02	1.30 (1.05–1.61)	1.3E-01	1.19 (0.95–1.48)	5.0E-02	1.27 (1.00–1.61)	3.4E-01	0.78 (0.47–1.30)
rs4823173	4.3E-03	1.35 (1.10–1.67)	5.9E-02	1.24 (0.99–1.55)	1.0E-02	1.38 (1.08–1.77)	6.0E-01	0.87 (0.51–1.47)
rs2076208	3.2E-02	1.36 (1.03–1.81)	6.3E-02	1.31 (0.98–1.75)	4.4E-02	1.34 (1.01–1.78)	4.3E-01	1.15 (0.81–1.65)
rs2294916	1.1E-02	1.31 (1.06–1.61)	1.1E-01	1.20 (0.96–1.49)	2.6E-02	1.32 (1.03–1.68)	3.6E-01	0.81 (0.51–1.28)
rs2294918	2.1E-01	1.22 (0.90–1.65)	5.7E-01	1.10 (0.80–1.50)	4.1E-01	1.15 (0.83–1.59)	3.2E-01	0.83 (0.56–1.21)
rs6006460	7.4E-02	0.24 (0.04–1.30)	1.4E-01	0.30 (0.06–1.63)	1.1E-01	0.28 (0.05–1.52)	2.3E-01	0.37 (0.07–2.03)
rs2294919	1.9E-02	0.71 (0.54–0.95)	4.3E-02	0.75 (0.56–0.99)	2.9E-02	0.73 (0.55–0.97)	4.5E-01	0.87 (0.61–1.24)

Supplementary Table 1c. ALD versus Healthy Association Tests.

					Alleles		Alt Frequency		Allelic Test	
rsID	Chromosome	Position	Category	LD	Ref	Alt	Control	Case	P	OR (95% CI)
rs4290029	1	222467263	Huang		C	G	0.69	0.68	2.9E-01	0.90 (0.74–1.09)
rs17740066	3	122582973	Huang		G	A	0.17	0.23	2.2E-01	1.17 (0.92–1.49)
rs62522600	8	103910885	Huang		G	A	0.02	0.02	4.2E-01	0.74 (0.40–1.36)
rs4986791	9	119515423	Huang		C	T	0.02	0.02	5.0E-01	1.34 (0.68–2.67)
rs886277	11	2396343	Huang		T	C	0.52	0.57	2.4E-03	1.32 (1.11–1.58)
rs2878771	12	48638660	Huang		G	C	0.12	0.08	4.6E-02	0.74 (0.56–0.98)
rs2290351	15	88175785	Huang		G	A	0.37	0.42	9.4E-03	1.28 (1.06–1.53)
rs9614194	22	42636398	LDtag	0.027	G	A	0.04	0.01	1.2E-03	0.40 (0.23–0.69)
rs5764023	22	42636551	LDtag	0.031	C	T	0.20	0.25	8.7E-02	1.22 (0.98–1.52)
rs4823168	22	42638868	LDtag	0.000	C	T	0.27	0.24	3.8E-02	0.80 (0.65–0.98)
rs929090	22	42645182	LDtag	0.010	A	G	0.55	0.56	9.9E-01	0.99 (0.82–1.20)
rs4823104	22	42652482	LDtag	0.021	A	G	0.05	0.03	2.5E-01	0.75 (0.49–1.17)
rs2076213	22	42654255	LDtag	0.070	T	G	0.19	0.26	8.3E-02	1.22 (0.98–1.53)
rs2076212	22	42654303	LDtag	0.037	G	T	0.10	0.07	2.1E-02	0.68 (0.49–0.93)
rs738407	22	42655288	LDtag	0.531	T	C	0.60	0.75	2.5E-07	1.67 (1.38–2.03)
rs139051	22	42656009	LDtag	0.473	A	G	0.38	0.23	9.2E-10	0.55 (0.46–0.67)
rs738409	22	42656060	Romeo	1.000	C	G	0.54	0.72	1.2E-08	1.69 (1.41–2.03)
rs738408	22	42656063	Romeo	0.988	C	T	0.54	0.72	2.4E-08	1.68 (1.40–2.02)
rs1883350	22	42659376	LDtag	0.776	T	C	0.56	0.73	1.9E-07	1.64 (1.36–1.97)
rs4823173	22	42660063	LDtag	0.833	G	A	0.49	0.69	5.4E-10	1.78 (1.48–2.13)
rs2076208	22	42662393	LDtag	0.273	G	C	0.84	0.88	6.0E-01	0.93 (0.73–1.19)
rs2294916	22	42672255	LDtag	0.791	T	G	0.49	0.67	6.0E-09	1.71 (1.43–2.05)
rs2294918	22	42673449	LDtag	0.352	A	G	0.76	0.87	2.9E-08	1.92 (1.53–2.42)
rs6006460	22	42673507	Romeo	0.008	G	T	0.00	0.00	5.6E-01	1.78 (0.48–6.60)
rs2294919	22	42673658	LDtag	0.279	C	T	0.17	0.13	8.6E-01	1.03 (0.81–1.30)

	Logistic Regression Likelihood Ratio Tests							
	No Ancestry Correction		Global Correction		Global & Local Correction		Condition on rs738409	
rsID	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)
rs4290029	9.6E-01	0.99 (0.79–1.25)	3.9E-01	1.11 (0.87–1.41)				
rs17740066	1.0E-01	1.27 (0.95–1.70)	2.9E-01	1.17 (0.87–1.57)				
rs62522600	2.8E-01	0.68 (0.33–1.38)	5.1E-01	0.79 (0.38–1.61)				
rs4986791	7.4E-01	1.14 (0.52–2.48)	4.7E-01	1.33 (0.60–2.94)				
rs886277	3.7E-02	1.26 (1.01–1.56)	1.3E-01	1.18 (0.95–1.47)				
rs2878771	3.3E-01	0.85 (0.61–1.18)	9.7E-01	1.01 (0.71–1.42)				
rs2290351	7.2E-02	1.23 (0.98–1.53)	4.1E-01	1.10 (0.87–1.39)				
rs9614194	4.5E-03	0.41 (0.22–0.77)	4.0E-02	0.52 (0.27–0.98)	6.8E-02	0.56 (0.29–1.06)	1.0E-01	0.59 (0.31–1.12)
rs5764023	7.3E-01	1.05 (0.80–1.38)	9.9E-01	1.00 (0.76–1.32)	9.3E-01	0.99 (0.75–1.30)	8.9E-01	0.98 (0.74–1.29)
rs4823168	2.7E-01	0.87 (0.68–1.11)	3.5E-01	0.89 (0.69–1.14)	5.8E-01	0.93 (0.72–1.20)	3.4E-01	0.88 (0.69–1.14)
rs929090	8.6E-01	0.98 (0.78–1.23)	5.1E-01	0.92 (0.73–1.17)	5.5E-01	0.93 (0.74–1.17)	4.7E-01	0.92 (0.72–1.16)
rs4823104	4.4E-01	0.82 (0.50–1.35)	9.9E-01	1.00 (0.60–1.65)	7.7E-01	1.08 (0.65–1.80)	8.8E-01	1.04 (0.62–1.73)
rs2076213	9.1E-01	1.02 (0.78–1.33)	4.7E-01	0.90 (0.68–1.19)	4.6E-01	0.90 (0.68–1.19)	2.7E-01	0.85 (0.64–1.13)
rs2076212	6.8E-02	0.71 (0.49–1.03)	2.1E-01	0.78 (0.53–1.14)	2.1E-01	0.78 (0.54–1.15)	3.3E-01	0.83 (0.56–1.22)
rs738407	1.9E-03	1.45 (1.15–1.84)	1.5E-01	1.20 (0.94–1.55)	3.4E-01	1.14 (0.87–1.49)	9.8E-01	1.00 (0.71–1.39)
rs139051	1.4E-03	0.69 (0.55–0.87)	7.4E-02	0.80 (0.63–1.02)	3.2E-01	0.87 (0.67–1.14)	4.3E-01	0.88 (0.63–1.22)
rs738409	8.4E-04	1.45 (1.16–1.80)	5.1E-02	1.26 (1.00–1.58)	1.9E-01	1.18 (0.92–1.51)		
rs738408	1.4E-03	1.43 (1.15–1.78)	7.7E-02	1.23 (0.98–1.56)	2.7E-01	1.15 (0.90–1.48)	7.1E-01	1.75 (0.10–31.2)
rs1883350	3.7E-03	1.37 (1.11–1.70)	1.3E-01	1.19 (0.95–1.49)	3.1E-01	1.13 (0.89–1.44)	7.0E-01	0.92 (0.59–1.42)
rs4823173	3.6E-04	1.47 (1.19–1.82)	5.0E-02	1.25 (1.00–1.57)	2.0E-01	1.18 (0.92–1.52)	6.1E-01	1.15 (0.67–1.96)
rs2076208	6.9E-01	0.94 (0.71–1.26)	5.7E-01	0.92 (0.69–1.23)	6.3E-01	0.93 (0.70–1.25)	5.4E-02	0.72 (0.51–1.01)
rs2294916	1.1E-03	1.42 (1.15–1.76)	9.2E-02	1.22 (0.97–1.53)	3.3E-01	1.13 (0.88–1.46)	9.8E-01	1.01 (0.61–1.67)
rs2294918	1.3E-03	1.59 (1.20–2.11)	4.3E-02	1.35 (1.01–1.82)	1.4E-01	1.26 (0.93–1.72)	1.8E-01	1.26 (0.89–1.78)
rs6006460	2.1E-01	3.31 (0.39–28.0)	9.2E-02	4.84 (0.56–41.7)	7.9E-02	5.08 (0.60–43.2)	6.3E-02	5.71 (0.65–50.0)
rs2294919	8.0E-01	1.04 (0.78–1.37)	5.9E-01	1.08 (0.81–1.44)	6.1E-01	1.08 (0.81–1.43)	5.5E-02	1.39 (0.99–1.94)

Supplementary Table 2. Akaike information criterion (AIC) for rs738409[G] model selection.

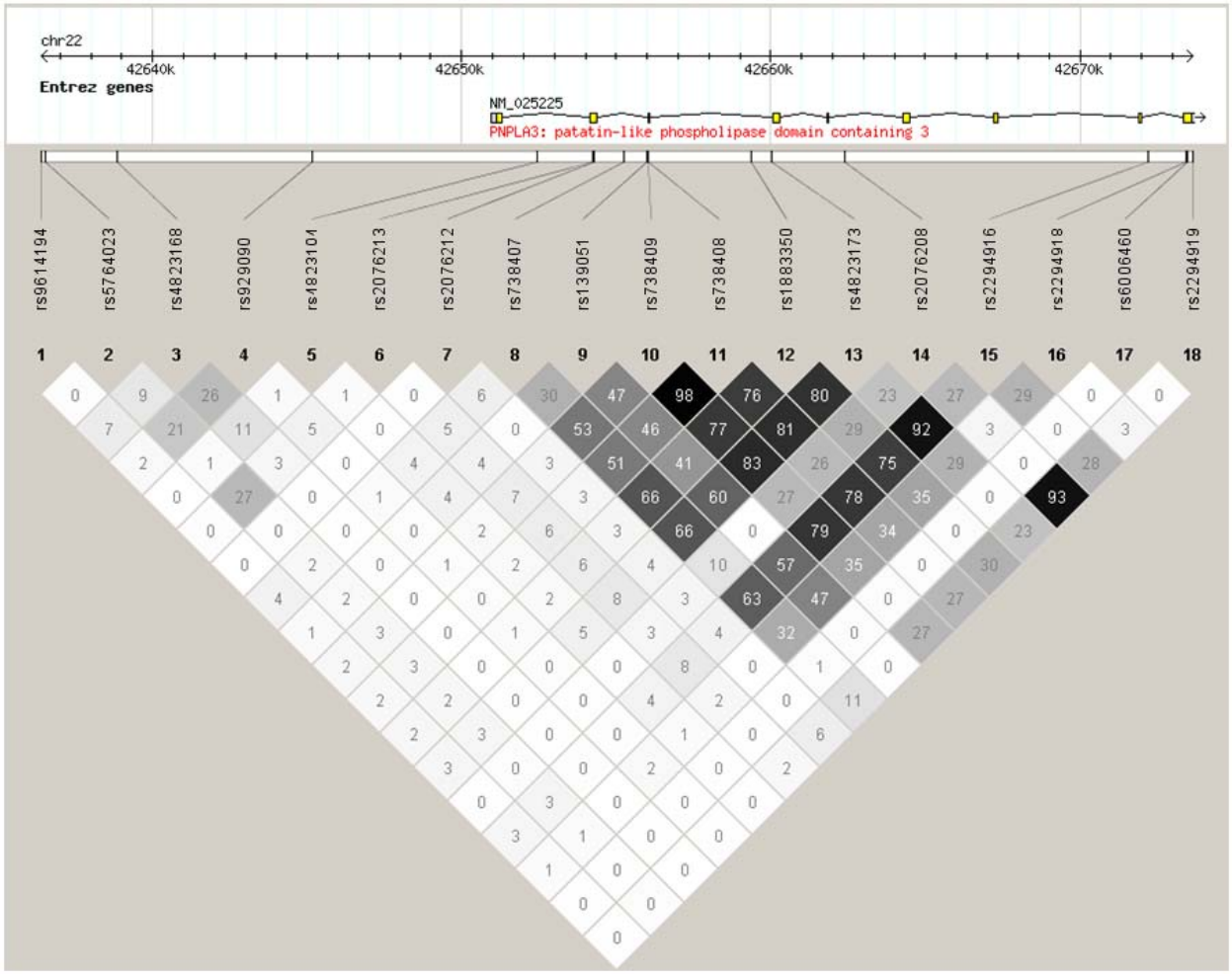
Model	Cirrhosis vs. Healthy	Cirrhosis vs. ALD
General 2-df	772.9	1091.6
Additive	770.9	1089.7
Dominant	779.9	1093.3
Recessive	774.8	1090.4

Supplementary Table 3. Haplotype association tests, for cirrhosis vs healthy status.

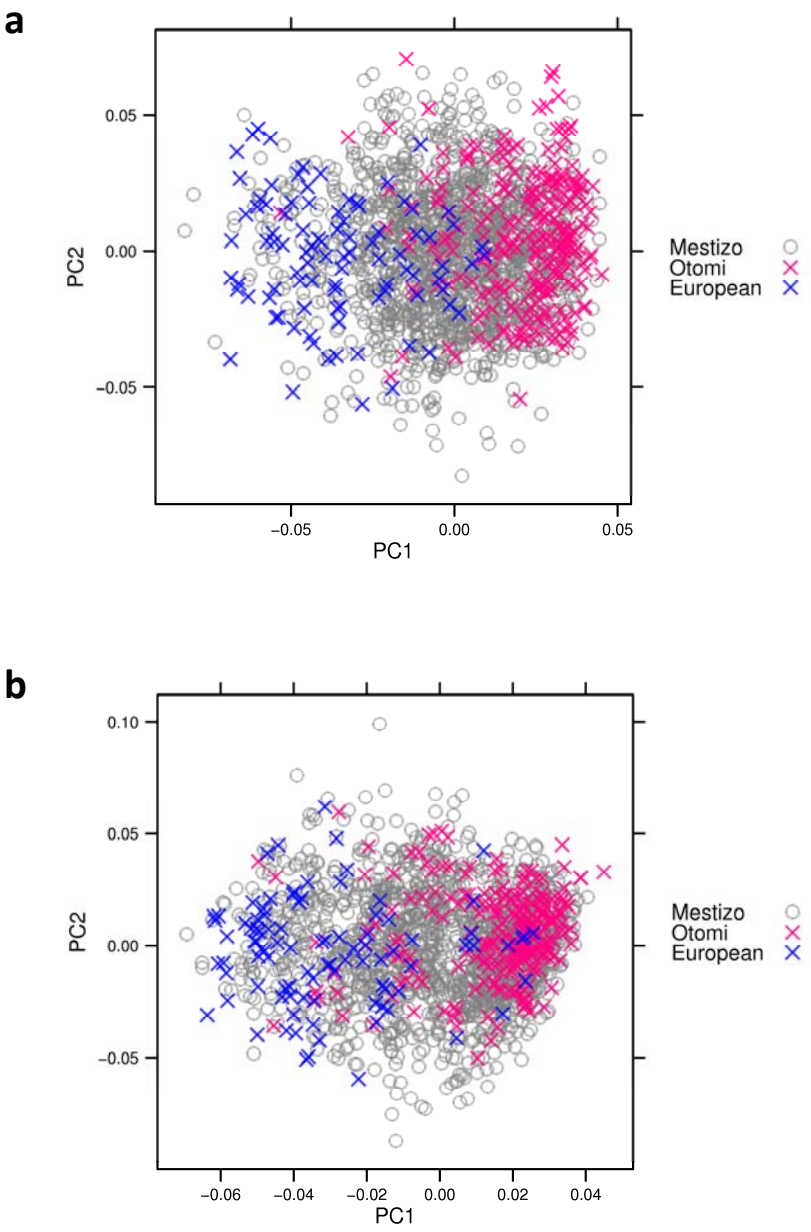
	Frequency		Pearson χ^2 test	Logistic Regression Corrected for Global PC1		
Haplotype	case	control	P	OR	95% CI	P
GCCAATGCAGTCACGGGC	0.2158	0.1672	0.0217	1.34	(1.01,1.80)	0.0439
GTCGAGGCAGTCACGGGC	0.1546	0.1115	0.0192	1.57	(1.09,2.27)	0.0133
GCTGATGCAGTCACGGGC	0.1369	0.0934	0.0122	1.77	(1.19,2.63)	0.0038
GCCAATGTGCCTGCTAGC	0.0664	0.1049	0.0085	0.55	(0.36,0.83)	0.0055
GTCGATGCAGTCACGGGC	0.0446	0.0311	0.2285	0.92	(0.50,1.67)	0.7828
GCCAAGGCAGTCACGGGC	0.0353	0.0295	0.6324	1.15	(0.61,2.17)	0.6662
GCCAATGTACCTGGTGGT	0.0332	0.0344	0.9908	1.20	(0.61,2.38)	0.5991
GCCGATGCAGTCACGGGC	0.0290	0.0279	0.9850	1.12	(0.54,2.30)	0.7576
GCTGGTGTGCCTGCTAGC	0.0145	0.0131	0.9908	1.11	(0.41,3.03)	0.8327
GCCGATTTACCTGGTGGT	0.0124	0.0246	0.1078	0.55	(0.24,1.22)	0.1447
GCTGATGTGCCTGCTAGC	0.0114	0.0311	0.0093	0.45	(0.19,1.08)	0.0658

Supplementary Table 4. Markers used for local ancestry assessment.

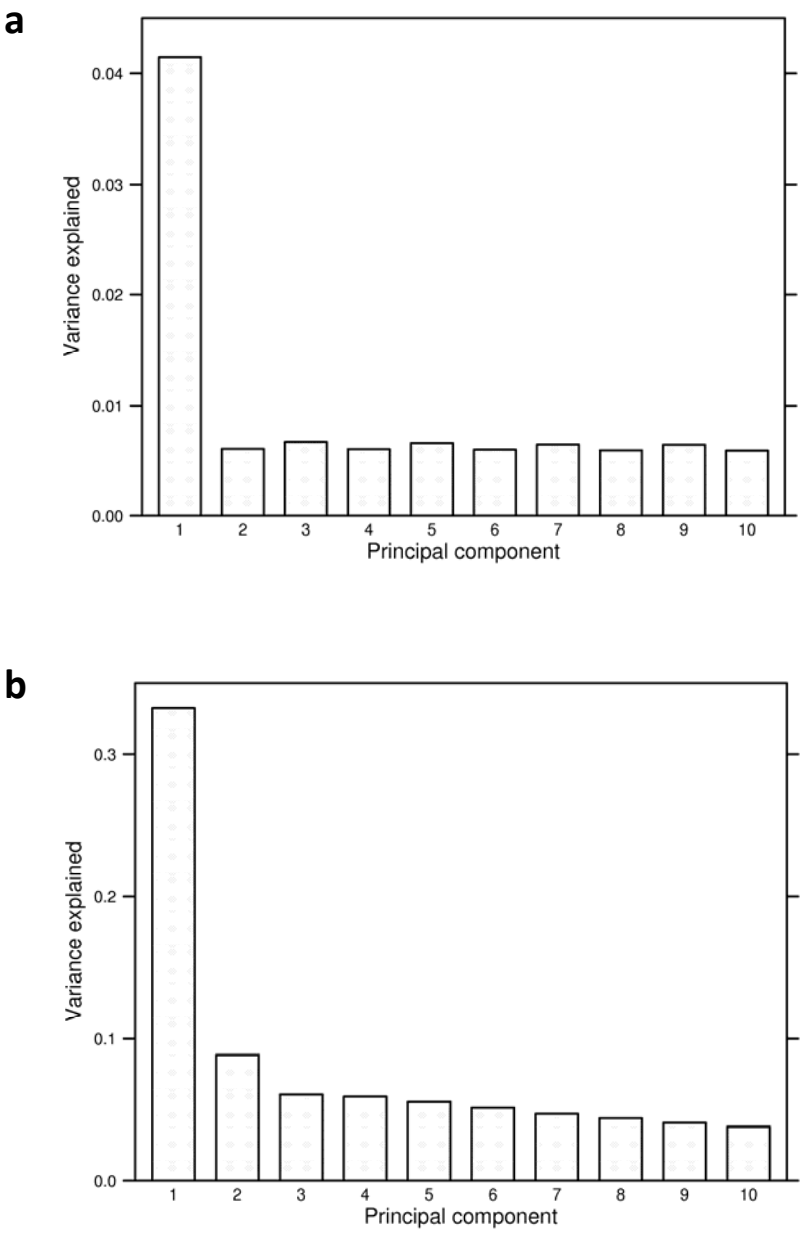
rsID	Chromosome	Position
rs739363	22	38090829
rs12484697	22	39066418
rs1984584	22	40014372
rs132793	22	40393627
rs7364180	22	40548802
rs2142695	22	40911192
rs2076158	22	41619682
rs733181	22	41984410
rs873724	22	43487691
rs961500	22	43717774
rs2283663	22	44320140
rs5767329	22	45473836
rs713999	22	46210776
rs5767830	22	46381466
rs5768714	22	47309834
rs132220	22	47463076



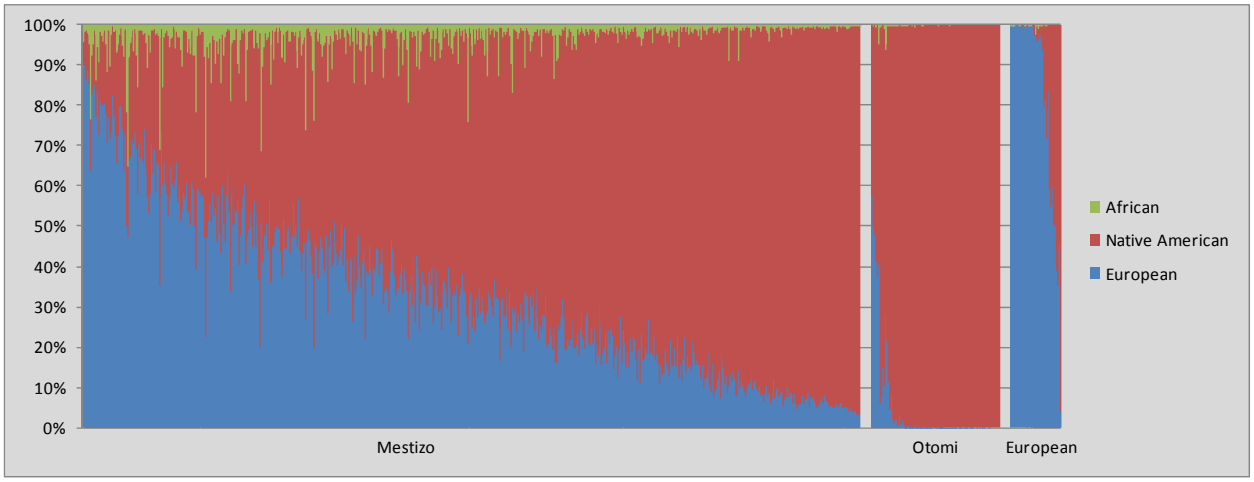
Supplementary Figure 1. Linkage disequilibrium of genotyped SNPs in the *PNPLA3* region, generated by Haploview[5]. Scores represent r^2 observed in our sample set.



Supplementary Figure 2. Distribution of (a) global and (b) local genetic ancestry along PC1 and PC2 estimated from PCA, colored by self-reported ancestry. In both analyses, PC1 measures Native American versus European ancestry; PC2 and higher components were not strongly correlated with self-reported ancestry.



Supplementary Figure 3. Variance components from (a) global and (b) local PCA. The first principal components account for substantially more variance than the higher order components.



Supplementary Figure 4. Global admixture proportions estimated from STRUCTURE. The individuals are sorted according to their global PC1 loadings, where larger loadings indicate higher Native American admixture proportions.